DOCUMENT RESUME

ED 168 470 IR 006 943

TITLE The Future of Organizing Knowledge. Papers Presented at the Texas A&M University Libraries Centennial Academic Assembly, September 24, 1976.
INSTITUTION Texas A and M Univ., College Station. Library.
PUB DATE 77
NOTE 56p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Automatic Indexing; Data Bases; *Futures (of Society); Information Needs; Information Processing; Information Retrieval; *Information Science; *Information Services; *Information Systems; Media Technology; Organization; *Technological Advancement; Telecommunication
IDENTIFIERS *Machine Readable Cataloging

ABSTRACT

The three papers presented at this assembly focused on communications technology and the information industry, organization of information, and development of the Machine-Readable Cataloging (MARC) system. "The Communications Revolution: America's Third Century Challenge" examines the growth of society from a pre- to a post-industrial society and the subsequent growth of the information industry and an information-based economy. While communications technologies can aid in systematically creating "information literacy" for all adults, it is also necessary to prepare individuals so that they may comprehend and apply communication capabilities and realize the potential of these technologies. "The Future of Organizing Knowledge: A Matter of Access" addressed the future of organizing information as the essential precursor of knowledge and wisdom, and claims that the ultimate objective of the information services is to provide access to information, thereby making information processing more efficient. The Machine-Readable Cataloging (MARC) system is examined in "The Many Facets of MARC," which covers the development of MARC, technical processing, files, acquisitions, and other significant aspects of the distribution service. (CWM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED168470

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE FUTURE OF ORGANIZING KNOWLEDGE

PAPERS PRESENTED AT THE TEXAS A&M UNIVERSITY LIBRARIES

CENTENNIAL ACADEMIC ASSEMBLY

September 24, 1976

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Richard L. Puckett

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Texas A&M University Libraries
College Station, Texas
1977

IR006943

TABLE OF CONTENTS

Introduction v

Program vii

The Speakers ix

The Communications Revolution: America's Third Century Challenge *by Lee G. Burchinal* 1

The Future of Organizing Knowledge: A Matter of Access *by Ralph E. O'Dette* 15

The Many Facets of MARC *by Lenore S. Maruyama* . . . 37

INTRODUCTION

Information, knowledge, and communications shaped in ways not yet completely fathomed the lives of Americans observing this nation's bicentennial and Texas A&M University's centennial. The contrasts between conditions in 1976 and those in 1876, a year in which this nation marked its centennial and the Agricultural and Mechanical College of Texas opened its doors, are so great as to tax the imagination.

The changes based upon the triad of information, knowledge and communications concern the Texas A&M University Libraries as well as libraries across America and throughout the world. The resources of human beings and of institutions are strained to meet the challenges of such changes. With these challenges in mind, the Texas A&M University Libraries faculty decided upon the topic of "The Future of Organizing Knowledge" for the Libraries' Centennial Academic Assembly. The series of Assemblies, arranged by the academic units of Texas A&M and presented through the Office of Continuing Education, helped to mark the University's centennial of teaching, research and public service. Various librarians assisted to make this Centennial Academic Assembly a successful event.

Upwards of one hundred participants came together in the Rudder Tower on Friday, September 24, 1976, to consider "The Future of Organizing Knowledge". In addition to Texas A&M faculty and students, representatives of the library schools of Texas and faculty from other university libraries shared in the program.

Those in attendance were privileged to hear new ideas and concepts, and it is in the spirit of making the substance of these papers more widely available that they are published as the Libraries' Miscellaneous Publication 16.

Henry L. Alsmeyer, Jr.
Coordinator, the Libraries'
Centennial Academic Assembly

MORNING PROGRAM

Presiding

Irene B. Hoadley
Director of Libraries
Texas A&M University

10:15 a.m.

Welcoming Remarks

John C. Calhoun, Jr.
Vice President for Academic Affairs
Texas A&M University

10:30 a.m.

The Communications Revolution: America's Third Century Challenge

Lee G. Burchinal
Director, Division of Science Information
National Science Foundation

AFTERNOON PROGRAM

Presiding

Henry L. Alsmeyer, Jr.
Associate Director of Libraries
Texas A&M University

1:15 p.m.

The Future of Organizing Knowledge: A Matter of Access

Ralph E. O'Dette
Senior Staff Advisor to the Director
Chemical Abstracts Service

2:45 p.m.

The Many Facets of MARC

Lenore S. Maruyama
MARC Development Office
Library of Congress

The Speakers

Dr. Lee G. Burchinal brought to the post of Director, Division of Science Information, National Science Foundation, a broad range of interests and accomplishments. The author of more than 150 scholarly publications, he is also a past president of the National Micrographics Association. His academic career includes service as a Research Professor at both Iowa State University and The Ohio State University.

Mr. Ralph E. O'Dette, the Senior Staff Advisor to the Director of Chemical Abstracts Service, served earlier with federal agencies and in industry. While with the National Bureau of Standards, he was engaged in their publications program, and during his service with the National Science Foundation he served as the Secretary to the Science Information Council.

A graduate in chemical engineering of the University of Virginia, O'Dette is a member of the American Chemical Society and the American Society for Information Science. He is chairman of the Advisory Committee on the World Inventory Data Base of the National Federation of Abstracting and Indexing Services and is a member of the Board of Directors of Information Science Abstracts.

Ms. Lenore S. Maruyama joined the Library of Congress upon receiving the Master of Arts degree in Library Science from the University of Wisconsin at Madison. Since 1967 she has been a member of the Library of Congress staff developing the MARC (Machine Readable Cataloging) system and serves as a projects supervisor. Her responsibilities include projects dealing with authority files, automated book catalog production, microforms, and data bases for the Register of Additional Locations and the National Union Catalog.

THE COMMUNICATIONS REVOLUTION:
AMERICA'S THIRD CENTURY CHALLENGE

By

Lee G. Burchinal

In this bicentennial year, we are particularly mindful of our revolutionary origins. The year of 1776 was one of political revolution. Since then the United States has undergone two other revolutions. Both are technological. Both have altered every aspect of our society. The first of these technological revolutions was represented by the conversion of United States society from a pre-industrial to an industrial society. The second is now in full swing and has been characterized by Dr. Daniel Bell of Harvard University as the emergence of the world's first post-industrial society.

The American colonies were a pre-industrial society. Before 1800, more than 90 percent of our population lived on farms. Life was a "game against nature," as Dr. Bell puts it, in which life was a struggle against the vicissitudes of the changing seasons. Power came from nature--water, and animal and human muscle. The strategic resource was raw materials--food, fiber, lumber and some minerals. Problem-solving was based on tradition, common sense, trial and error, and experience.

By the mid-19th century, the United States was well established as an industrial society. Population became concentrated near factories. Cities grew. Life for most became controlled by production of things. We used created energy--coal, oil, gas, electricity, and more recently, nuclear power--to add to our power to alter the landscape. Financial capital and labor became the strategic resources of production. Technology, which had been limited to the creation of artisans, switched to machine-generated products and rested on the skills of engineers and skilled and semi-skilled workers. Problem-solving shifted to empiricism and experimentation.

This was also the era of great mechanical inventiveness. Innovations, however, came from the brains of talented thinkers who generally remained ignorant of the discoveries and theories of contemporary scientists. For example, Alexander Graham Bell, our much heralded inventor of the telephone, knew nothing of the work of James Maxwell on electromagnetism. Nevertheless, invention piled on top of invention. Technological development snowballed as the nation pursued unlimited economic growth.

Now, however, as we analyze our society, we see that another massive transformation is underway. We are passing from an industrial into a post-industrial society. The strategic resources of a post-industrial society are information and knowledge. By information, I refer to any type of data or facts about anything. Included are data about population, raw materials, manmade products, the environment, and all types of economic, commercial, and governmental transactions. Included are the data required for planning and managing economic and governmental activities and for making numerous personal choices. In addition, there is theoretical knowledge derived from scientific research about the workings of nature. Scientific knowledge about nature forms the basis of major new industries. Scientists, engineers, technicians and professional occupations supply the critical skills in our post-industrial

society. This trend became particularly evident after World War I. Before that time an American scientist with a Ph.D. rarely worked outside the university. Following the war, however, chemical engineering became a widespread profession in the United States, and chemistry produced remarkable results for industry.

Chemistry was not the only field of growing scientific and technological progress. Physicists began to open up new areas of progress in electricity and, later, electronics. In 1909, the General Electric Company recruited a promising young physicist, set him up in a laboratory, and permitted him to do basic research. That industrial researcher was Irving Langmuir, later to become a Nobel prize winner. In 1916, the Westinghouse Laboratory also hired a Ph.D. who was to become a world-renowned scientist, Arthur H. Compton. Other major industrial companies followed suit, set up laboratories for both applied and basic research, and hired scientists during the period between the two World Wars.

During World War II the present and growing values of electronics were forecast by the use of radar and the early computers. This was also the time when the seeds of the space age were sown and the greatest single example of scientific power--the release of atomic energy--was demonstrated. Of immense importance, of course, has been the invention of the transistor by Bardeen, Shockley, and Brattain, and the development of semiconductors to create the microelectronics industry. More recently, basic research on the possibility of passing a molecular beam through an optical field led to the creation of a maser and then a laser which, in turn, gave rise to the hologram. Applications of these discoveries abound, but nowhere are they more evident than in the field of information and data handling. Computers, micrographics, and telecommunications have become driving forces in our information-based society.

Emergence of the post-industrial society does not mean that agricultural and industrial activities have disappeared or have become unimportant. Quite the contrary. Because agriculture has become so efficient, largely due to science and technology, only 4 percent of our population now produce a substantial amount of our food, plus that needed by a large part of the rest of the world. Like agriculture, industry depends on scientific knowledge for its new growth. At the same time, industrial and commercial transactions increasingly depend on information and data processing. Examples include automated factories; reservation services for airlines and hotels; nationwide inventory control and sales records; and national and international news, entertainment, and communication services. These and many other industrial- and service-based jobs are information processing jobs, regardless of the goods or services involved. In addition, many other occupations, including research, development, teaching and various other professional jobs, can be described as information generating, processing or using activities. Collectively, they constitute the information industry. Just as agriculture is the base of the pre-industrial society, and

fabrication such is dominant in an industrial society, information or knowledge occupations characterize the post-industrial society. This conception may be difficult to grasp. We are attuned to physical representation of industrial society, such as factories and transportation systems.

In its early stages, the information industry was also highly visible. It began as the graphic arts industry. While dramatically different from the hand-written manuscripts of previous ages, printed products were still physical items. They were manufactured (printed), shipped and stored just like the things generated by factories. But with advances in printing technology, democratic values, science, and public education, the volume of printed material increased annually. The curve remains ever upward.

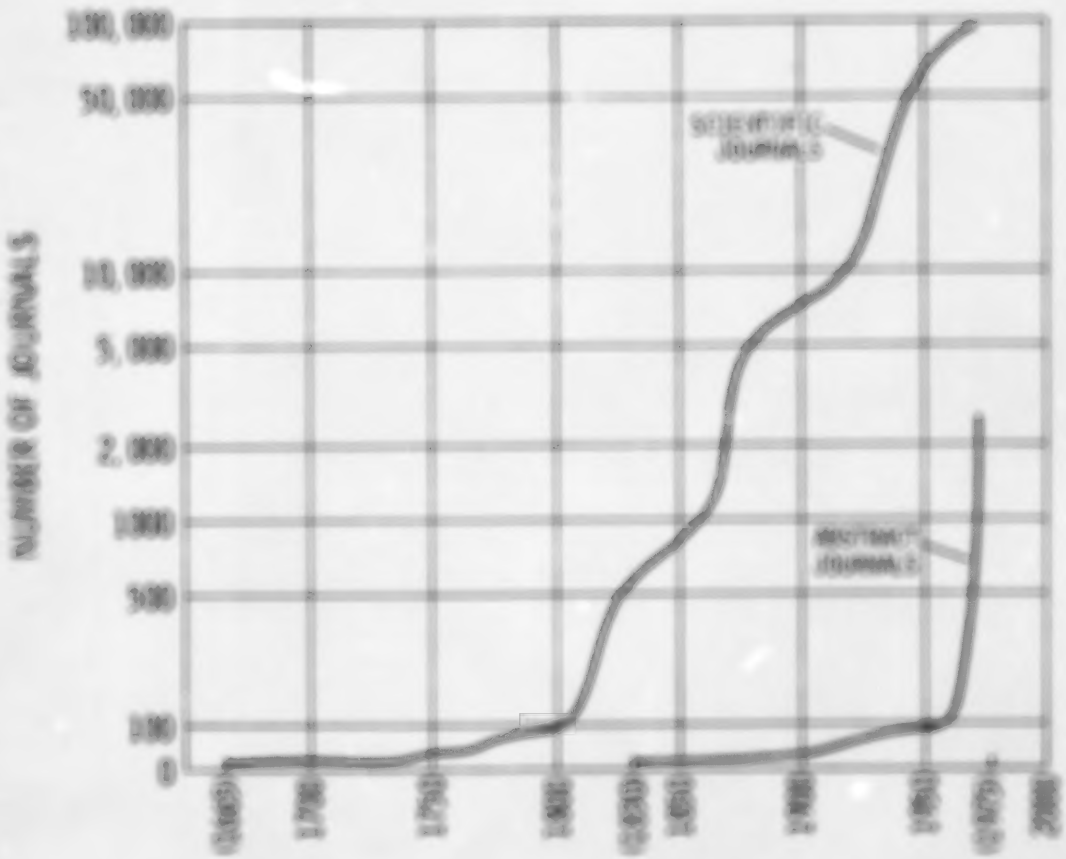
Library holdings from several points in time illustrate the long-term information--if not knowledge--explosion. Just before the invention of the printing press, a highly educated man's library--such as Chaucer's--might contain as many as twenty manuscripts. These he read and studied extensively. The first non-theological library in Colonial America, organized in 1731 under the leadership of Benjamin Franklin, initially consisted of 86 titles. Ten years later, it had grown to an impressive collection of 375 volumes.

A few years later and a little further south, another library was building along the Potomac. Its collection of 3,000 volumes went up in smoke in 1814 when the British burned parts of Washington. The next year, Thomas Jefferson provided the Library of Congress a second start with the sale of his collection of 6,000 volumes. Since then the growth of the Library of Congress has mirrored the ever-increasing worldwide production of literature. By 1900 the collection numbered 850,000 volumes. At the outset of World War II, the collection stood at almost 6.4 million volumes. And today it exceeds 70 million pieces of information.

Numerical increase is only one dimension of the "information explosion." Diversity of form is another. Library holdings illustrate this point as well. Today the Library of Congress--and other major libraries--contain books, journals, newspapers, technical reports, manuscripts, maps, microfiche and microfilm, motion pictures, video tapes, recordings, photographs, prints, posters, and computer tapes, disks and other forms of electronic storage.

Growth in the scientific and technological literature represents another indicator of the increase in the information base of our society. The first scientific journals, one French and the other British, go back to the mid-seventeenth century. Since then, the scientific literature has been doubling about every ten to fifteen years. As shown in Figure 1, over 100,000 journals have been published somewhere, some time: today, there are approximately 50,000 current titles in some sixty languages. About two million new articles or items are added annually to

WORLDWIDE GROWTH OF SCIENTIFIC AND ABSTRACT JOURNALS



the available scientific and technical literature. And such a growth, it is not surprising that the literature searching tools should develop as well. The first of these changes [illegible] in the early 1970s, and there are now two thousand scientific and technical abstracting/indexing publications.

During this period of tremendous growth in scientific and technical (STM) publication, there was relatively little change in the methods used. All were printed publications. But, beginning just about ten years ago, a remarkable change took [illegible]. Today most new handbook abstract [illegible] data bases equivalents are available in machine-readable form. The science of the computer file [illegible] of computer searching of STM data bases. This file first offered [illegible] searching five computer terminals located at the researchers in 1965. It began with one data base. By 1971, eleven separate files were offered. Today over forty data bases can be searched. The number of records available for online searching continues to grow. Again, the expansion of this file is [illegible]: from 400,000 in 1965 to 1.2 million in 1971 to 3.5 million in 1975. Records [illegible] several to about everything else in the science, some are [illegible]. Only an index of 100 for 1975 costs for online searching, 1971 costs had dropped to $1, and 1975 costs to $5. So as will see later, further decreases can be confidently expected.

Conversion of printed abstract journals to electronic form also [illegible] the available value of the information industry. Electronic means for recording, distributing and processing literature and data will [illegible] the same in the late 1980s. In contrast to use of printed journals and reports, we will only have highly [illegible] a terminal of information which have remote location. Steps to publication will be separate tape from word-processing machines and other forms of electronic storage.

The available value of the information industry adds to the difficulty of understanding its significance. The key to the answer is [illegible] each of the behavioral base of society. It is all around us. It is represented by factories, transportation facilities and products. The information society equivalent to both transportation systems are electronic communication systems. These include voice and data communication links, whether by voice, microwave or satellite [illegible].

But the transmission remains secure. Likewise, the occupational shift to information-oriented jobs has been associated in nearly 40 percent, but later and significant shifts have occurred. There are two ways of measuring the trend toward the dominance of the information industry in our society. One consists of examining trends in the portion of the gross national product (GNP) [illegible] to different economic sectors. The other focuses on the percentage of the labor force in various major occupational groupings.

KNOWLEDGE INDUSTRY AS A PERCENT OF GNP

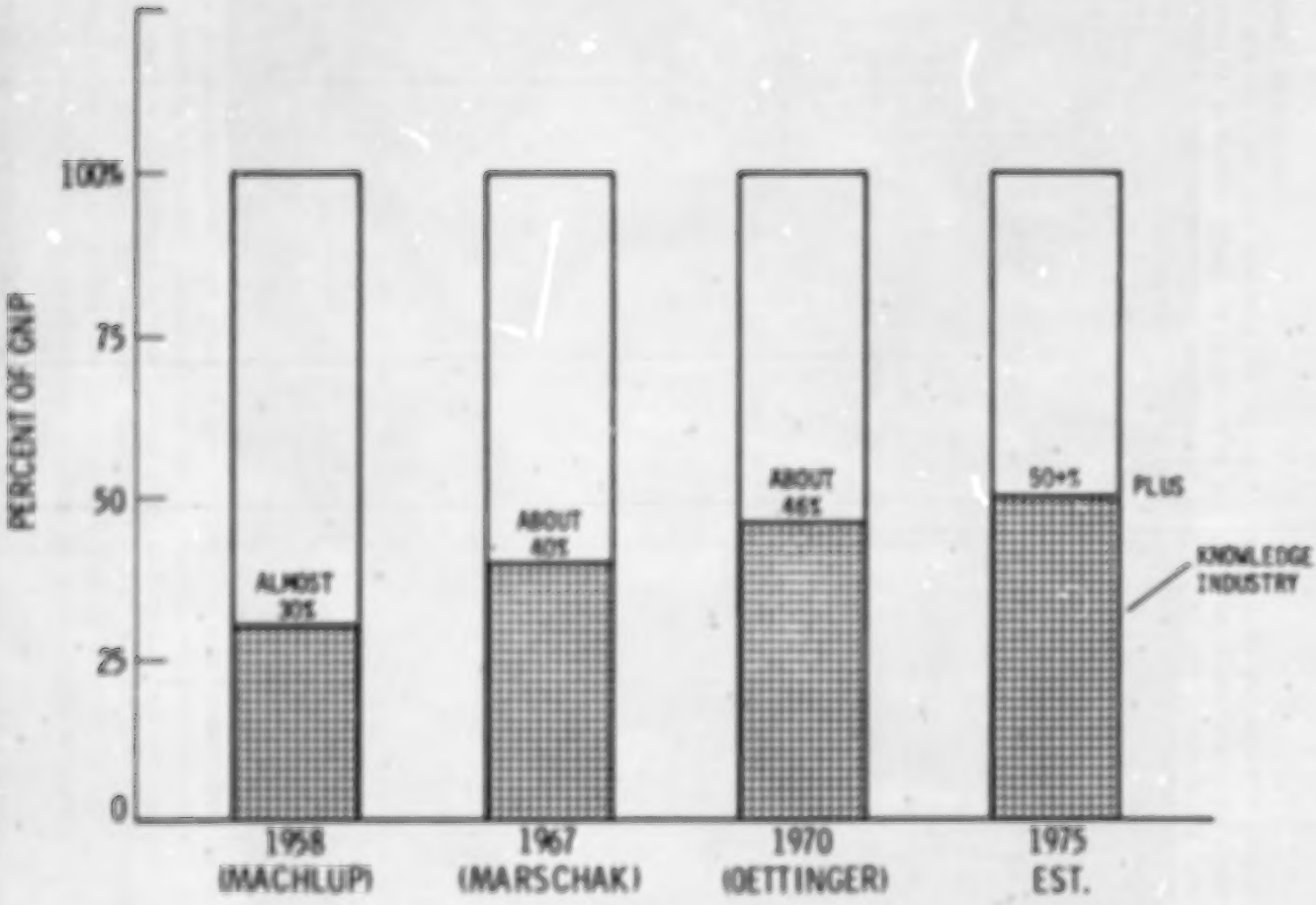


Figure 2

Let us look first at the GNP data. Several analysts have tried to estimate what proportion of the GNP is accounted for by activities concerned with generating, processing, disseminating, and using information and knowledge (Figure 2). The first estimate was made in 1958 by Professor Machlup in his book, The Production and Distribution of Knowledge. He put the figure at nearly 30 percent. In 1967, Professor Marschak reworked available data, and raised the estimate to almost 40 percent. Three years later, Dr. Oettinger further upped the figure to 46 percent. Today, experts believe that over half of the GNP is contributed by the information industry.

Analysis of labor force data leads to the same conclusion. Here we are indebted to analyses developed by Marc Porat and Dr. Edward Parker from Stanford University. Their definition of information occupations include those jobs in which:

> Information is a final product. This includes the newspaper and printing industry; broadcasting and motion picture; telephone and telegraph; libraries and postal service.
>
> Information is a major intermediate product. Included here are financial institutions; real estate; legal; advertising; education; engineering and architecture; accounting and auditing; government administration and commercial research and development (R&D).
>
> Information industry. Occupational fields in this case include computer and communications; scientific instruments; and office, computing, and accounting equipment manufacture sales and service.

Remaining occupations are classified as being in non-information industries.

Using these definitions, the Stanford researchers traced trends in U.S. information and non-information fields since 1860 (Figure 3). Information occupations started at about 5 percent of the work force in 1860 and remained approximately 10-15 percent until 1950. Immediately following World War II, occupations in the information sector exploded in size and grew rapidly. During the present decade (1970-1980), the information occupations emerged as the largest group in the work force. It is unlikely that this trend will reverse.

Porat and Parker also undertook a more detailed analysis. They divided the work force into four major independent categories: agriculture, industry, services and information. Independence was achieved by separating information jobs in agribusiness, such as researchers, sales persons, accountants, secretaries, and other information jobs, from other agricultural jobs. The industry classification

U.S. LABOR FORCE
2 SECTOR AGGREGATION
(LOW ESTIMATE ONLY)

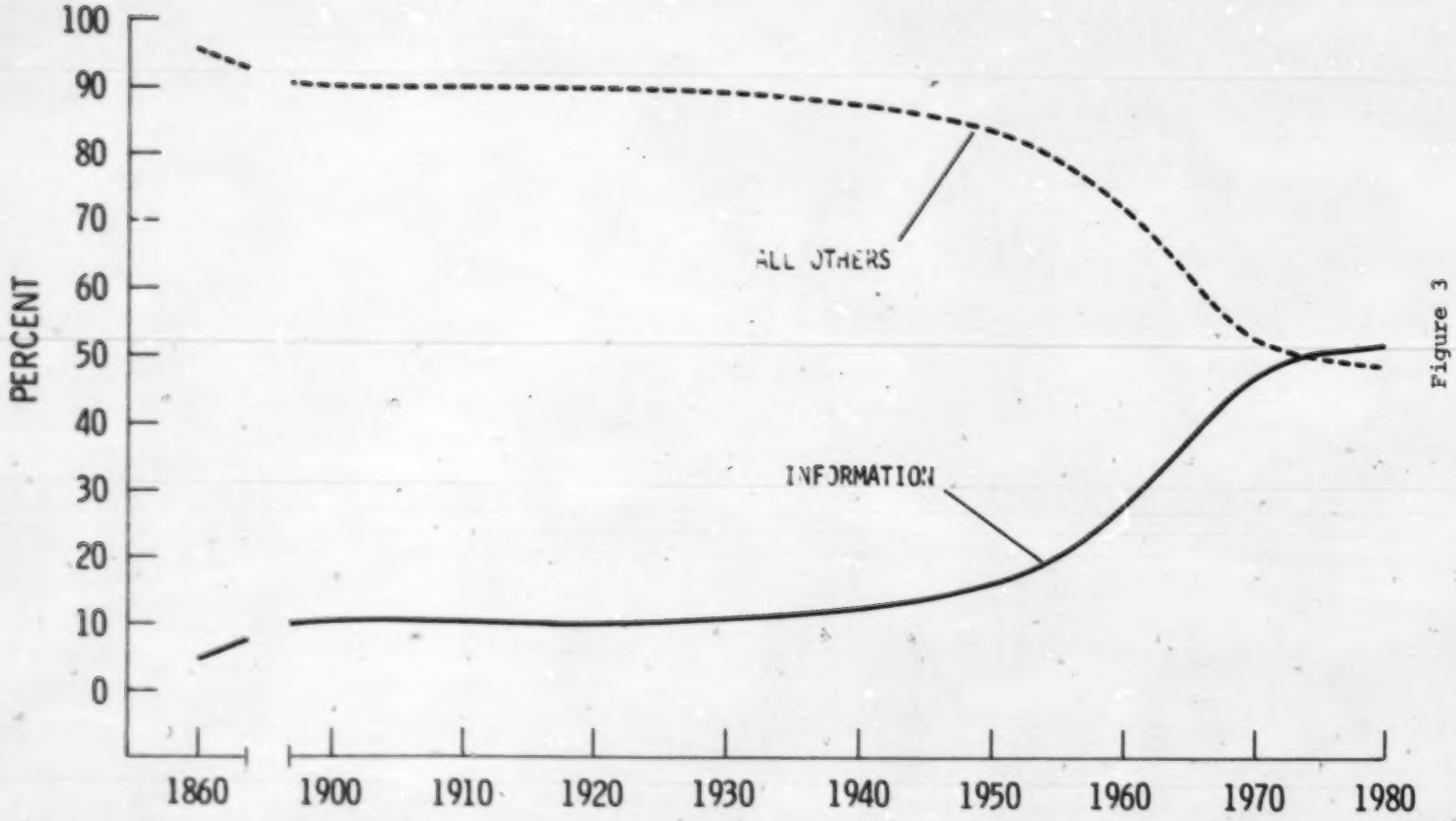


Figure 3

included blue-collar plus all other jobs not classified elsewhere. Service cluster includes domestic, repairmen and public service workers engaged in non-information work (policemen, firemen, or sanitation workers). While more illuminating, the results are the same as before.

But, with these distinctions, we can identify several separate trends (Figure 4). Prior to 1860, the largest labor group in the United States was the agricultural, comprising 36 percent of the work force. By 1900, industry had begun to take over, absorbing workers from the countryside into the urban factories. At this time, 20 percent of the work force was in services, largely personal and domestic, while information workers amounted to only 10 percent of the labor force. As the agriculture sector contracted and receded to its present 4 percent of the work force, the industrial sector grew rapidly. Between 1935 and 1955, the United States was predominantly an industrial society, peaking in 1950 with about 65 percent of the work force.

Around 1945, we had almost equal proportions of agriculture, services and information workers, around 12 percent of the work force. Services and agriculture had been receding slowly, and information had been increasing. After this period, however, the information sector grew very quickly in proportion, not by pulling people away from the industry sector (whose membership in absolute terms grew somewhat), but by capturing most of the new entrants into the labor pool. In the late 1960s the information sector surpassed a quickly dwindling industrial sector, and is now by far the largest single group. These curves closely mirror Daniel Bell's thesis that the United States has progressed through three stages--from agriculture to industrial to post-industrial; and that we are now firmly embarked on the third era.

Trends toward an information-based economy are well established. A number of driving forces stand behind these trends. One is the search for more efficient and cost/effective ways of producing goods and services. Applied research and development expenditures are an indication of this force. Another is expenditures in basic research in order not to deplete the stock of theoretical knowledge. Further, R&D expenditures have shifted to the communications field. In recent years, aerospace dominated United States R&D. In 1973, R&D on electronics and telecommunications took over the lead. From this research will come future information-processing and telecommunication technologies, industries, systems and jobs.

Another driving force is the cost/effective attractiveness of electronic-based operations. To illustrate:

> Computer costs are declining exponentially, while cost/performance is increasing similarly. This includes costs and performance for main frames, mini-computers, micro-processors, and related storage devices.

U.S. LABOR FORCE
FOUR SECTOR AGGREGATION
U.S. LABOR FORCE

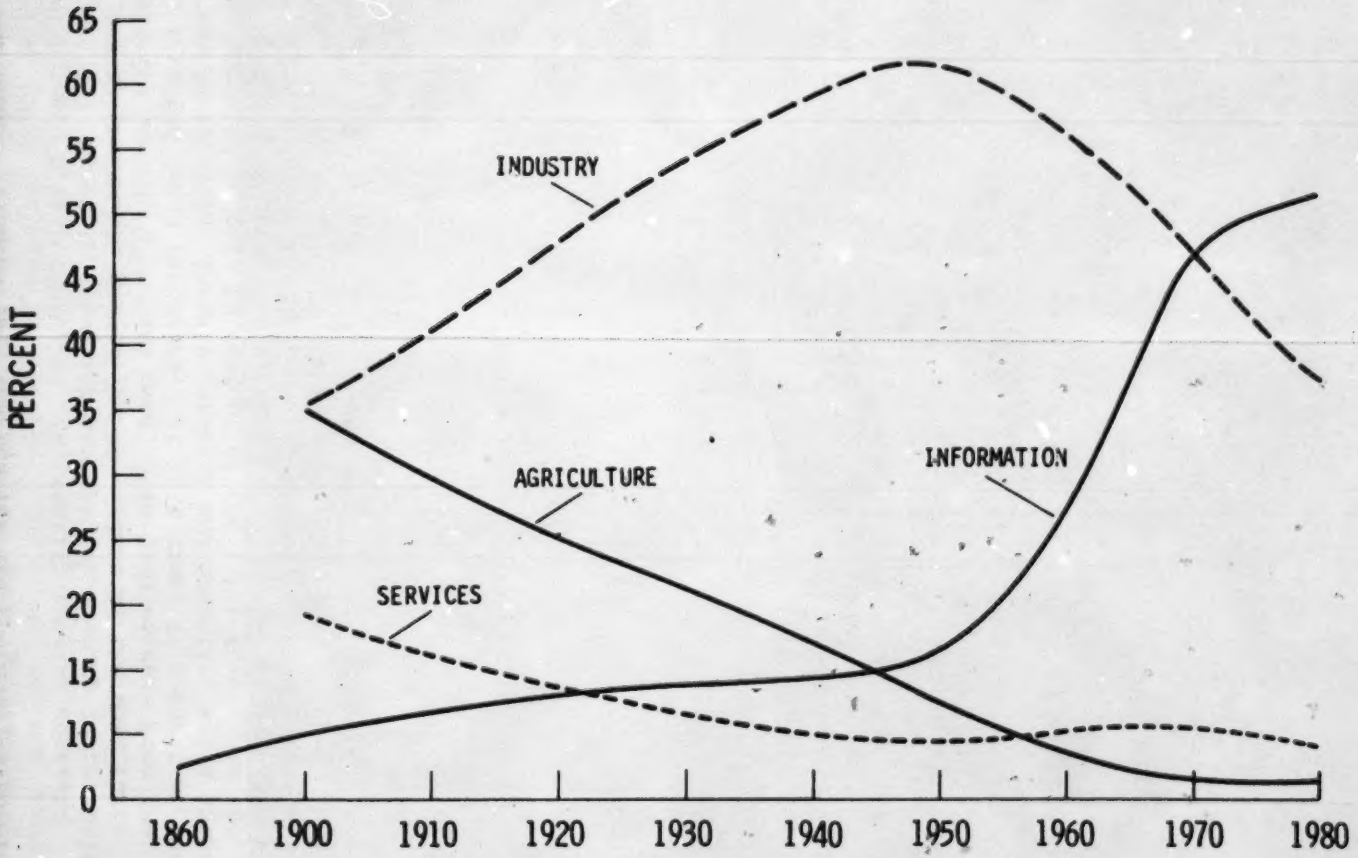


Figure 4

> Costs for terminals are also declining while performance is constantly improving.
>
> Telecommunications, which have remained high relative to computing operations, are now declining and will drop drastically in the future.

In contrast, conventional communication, based on the printed word and postal services, will continue to increase in cost.

Further impetus for accelerated development and use of communication technologies comes from our desires to conserve energy. In contrast to moving people, which represents high energy consumption, communications based on audio and video transmission or computer conferencing can allow communications and decision-making, without the costs of transportation and high energy consumption.

The net result will be to accelerate communications via electronic networks. As a result, more jobs and personal experience--banking, purchasing, and communications with friends and work associates--will be based on use of terminals.

Thus, the electronic media and the post-industrial society are extricably intertwined. Universities have recognized this fact. Engineering, science, business administration and other curricula include instruction in computer operations. Computer, library, and information science instruction is in healthy ferment. Individuals initially educated in one of these fields move easily to the other and, further, to a wide range of jobs in the public and private sectors. As welcome and important as these changes are, two further and even more significant improvements, I believe, are necessary if we are to reap the full benefits of the coming communications revolution.

First, we should set about systematically to create "information literacy" for all adults in the nation, so each can function effectively in our emerging society. Second, we should begin preparing individuals who will be able to comprehend and apply our new communication capabilities creatively across the spectrum of society's needs so that we fully benefit from the potential of our new technologies.

Information literacy goes beyond conventional literacy--the ability to read and write in acceptable--or at least passable--English. Regardless of how successful we have been with conventional literacy, we need to look ahead to achieving a new kind of national literacy--information literacy. To be information literate requires a new set of skills. These include how to efficiently and effectively locate and use information needed for problem-solving and decision-making. Such skills have wide applicability for occupational as well as personal activities. Part of such competency includes comfortable use of a computer terminal for sifting through available information from various data banks to select useful data for resolving the problem at hand.

No college graduate today should be without such skills. Consider our 1977 graduates. By the year 2,000, they will be in their mid-careers. By then, many will have substantial management responsibilities. Management roles will be greatly influenced, if not dominated, by automated communications. Even by 1990, only thirteen years out, all indicators point to substantially electronic means of distributing information. Freshmen entering the labor market in less that ten years face similar prospects. Job performance, including management competence, will depend upon information literacy as well as competency in one's chosen field.

Also, if our present forecasts are correct, computer-terminal devices will be common in the home within a decade. From such devices we will do our banking, shopping, plan vacations, and communicate with friends. We will have our choice of voice, digital, graphic, or pictorial forms of presentation. All family members will use such devices, just as preschool children today master the telephone, record and tape players, and television sets.

As these technologies become more common, elementary schools will take over the responsibility for creating information literate citizens. Universities, however, can ill afford to wait. Also university experience as in so many fields, can become the basis for subsequent school programs.

Second, we need to give attention to preparing individuals who can anticipate and develop appropriate data teleprocessing and related telecommunications systems for whatever needs to be done--in managing economic activities or providing educational and community services. The only real limit to the development of the automated systems is the shortage of people who can comprehend them, develop them, change them, and anticipate the consequences of their operations. As such systems become more sophisticated, people with the necessary perspective and understanding become scarcer. We may be deceived by the ease with which we can train technicians to do specialized tasks required by computers and information processing networks. What is critically needed are people who understand these new systems and how they can be applied to economic, health, educational, and community requirements. Individuals with these skills command a premium in the emerging post-industrial society. Universities with foresight and resolve can be in the vanguard in meeting the nation's future needs in this critical area.

As we enter our third century, the post-industrial society and its information and communication base present numerous challenges. I am confident we will apply these new resources for the benefit of all our citizens. The communications area allows us to focus three of our outstanding and unique capabilities. These are our confidence in ourselves and optimism about what we can accomplish; our ability to sustain technological innovation; and our management expertise.

Here is a challenge fitting for our third century....

THE FUTURE OF ORGANIZING KNOWLEDGE:

A MATTER OF ACCESS

By

Ralph E. O'Dette

The title of this centennial symposium is stimulating, challenging, and difficult to deal with. It is often more fun to speculate about the future than it is to confront the present or explain the past. It is comforting to know that no one will bother to look back in years to come to verify the accuracy of my foresight. But the future is also a serious matter, and I will try to deal with it seriously. By "seriously" I do not mean solemnly; I will deal with a view of the future that I am trying hard to help bring about, and I think it will be a bright future, at least with respect to the topic of our symposium.

The "future of organizing knowledge" has several connotations of importance. One profound and difficult connotation is that the intellectual structure of knowledge may or should or perhaps even will change. Another inference that can be drawn is that the organizers of this symposium, recognizing a dichotomy between information and knowledge, are stressing their concern for knowledge. At a more mundane level, the phrase suggests concern for how the human race will manage the record of what it knows so that we will not repeat ourselves unless we want to.

I believe that a centennial celebration demands a philosophic discourse, but I am an engineer, and an amateur philosopher at best. Regardless, I was asked to deal with the future of scientific publishing, using the example of Chemical Abstracts Service, and so I will explore the mundane connotation. I will deal with facts and their immediate implications, but I hope you will forgive me if I strive for an occasional insight--perhaps even a deep insight.

Information, knowledge, and wisdom form a hierarchy. Wisdom is most important, with knowledge next; however, neither can exist unless they are preceded by information. Conversion of information to knowledge or wisdom is the task of the receiver of the information.

My assignment today requires me to brush that profound problem aside, also. I will concentrate on the future of organizing information as the essential precursor of knowledge and wisdom. The title of my remarks, "A Matter of Access", simply says that in dealing with information, what matters most is that the seeker find what he needs to know. I will talk about a number of aspects of helping users of information to find what they need to know. I will strive for generality, but my illustrations will come principally from Chemical Abstracts Service (CAS).

In discussing CAS, I will first deal with the role of the accessing service in the de facto international information system. Then I will trace our adoption of modern information handling technology, and, finally, I will suggest what I believe our work means for the future. It is certainly not my purpose to teach you how to run an accessing service, to sell you a subscription to CA, nor even to teach you how to use the subscription you may have. You have my permission to forget every detail I describe if you promise to try to remember the principle each is meant to illustrate.

Figure 1 is at least ten years old, but its meaning and significance appear to be eternal. The information user is, or feels he is, always in the middle, and he is correct. Not before the advent of the ideal information system, if even then, will the user be able to escape dealing with multiple sources of information according to the type of information he wants. Each of the boxes indicates an established kind of information activity. In real life there is some overlap among the boxes, but the main function of each is basically different from that of the others.

It may require genetic engineering to cure scientists of the habit of publishing papers. Some of the characteristics of the primary journal are changing, but whatever new media or techniques evolve, there must be some form of first disclosure of new information, and that is what is meant by the "original literature" box.

Accessing services now collect, concentrate, and organize large quantities of original literature to guide the user to that part in which he is interested but which he may not have encountered directly. How the user will discover and gain access to this information in the future will be different. As the nature of the original record changes, the accessing services will accommodate to the changes.

Community information centers are, specifically, computer-based search centers, and they are growing in importance in the information universe.

Data centers are sources of, for the most part, numerical information, such as about the biological, chemical or physical properties of substances. The management of what might be defined broadly as measurement results is a difficult and costly problem that nevertheless must be solved. I think data management is the least understood of the basic functions in my information universe.

Libraries obviously play a unique and critical role in the information universe. They sometimes--but not as often as one might expect--operate a community information center. Both the users and the accessing services depend on libraries to have or to find needed original literature.

I said that in real life there is overlap among the boxes in Figure 1; however, the separations between the boxes are more truth than one would wish for. The extent of meaningful cooperation and coordination among the basic functions of the information transfer community is pitifully small. Each suffers today from what I call the railroad syndrome: each tends to have a myopic view of itself and the business it is in. Perhaps if the railroads had decided in the beginning that they were in the transportation business rather than the railroad business, the industry might have become more far-sighted and flexible, developing in a different relation to automobiles, trucks, airplanes, and perhaps even telecommunications. In a broader context, the railroads themselves might have stayed healthy.

THE INFORMATION UNIVERSE

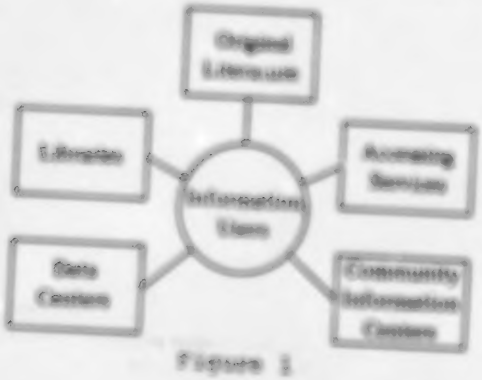


Figure 1

SOURCES OF CHEMICAL INFORMATION

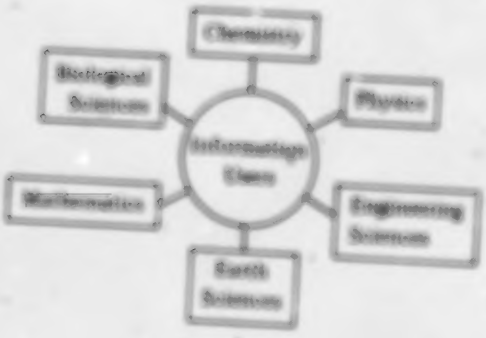


Figure 2

A genuine acceptance of interdependence is essential, I feel, for each of these steps in the information universe to do best and most efficiently what is expected of it, and so that all, in combination, supply superior service to the user. The producers of original literature must stop considering themselves simply as primary journal publishers even if that remains their principal activity for centuries to come. The classical primary journal is only one way to create a useful record of original information. The screening services must take more and more to that identification option that call themselves "secondary service" or the still narrower "abstracting and indexing journal". The screening services need a better liaison with libraries to help support the libraries' role as organizers of the bibliographic record and as providers of source documents to users. Data centers, screening services, and original literature sources must work much more closely to help solve the data management problem. That problem, viewed across the full spectrum of science and technology, is almost incomprehensibly large and complex. It absolutely demands team work. Information centers and libraries need to be more on speaking terms than is usual today, the centers by acknowledging their users' frequent need for source documents, and the libraries by recognizing the computer-based center as a reference function.

Libraries might also admit that the screening services provide organized subject access to the records, conference proceedings, and even monographs that the libraries hold. Libraries must take care not to interpret their truly unique role to mean that they must occupy the center of the universe.

I am sure we could increase this list almost endlessly. A spokesman for each function could easily produce a long list of desired changes for the others to make. I am also sure that the future of organizing knowledge will be much brighter if Figure 1 becomes a chain rather than a group of isolated boxes.

Figure 2 poses a structurally similar argument about relations among disciplines. This is a deep and difficult problem that I believe has for the most part been dealt with superficially. In Figure 2, the question is posed in terms of the needs of an information user for chemical information. A typical chemist, if there is such a person, probably finds much of the information he needs in packages such as journals and books that have the word "chemistry" on their titles. But there are times when he wants information that the producers have labelled differently, as biology or physics, for example.

Figure 2 is grossly oversimplified since there are only a few basic disciplines on the chart. Even those disciplines are not discrete but quite routinely overlap. Physical chemistry and chemical physics are obvious examples. Biochemistry is a much larger example, an well established that it has the character of a distinct discipline itself. Definition becomes more difficult with such topics as energy and ecology which are of great current interest. These topics are highly multi-

Encyclopedia and could not be fit into an alphabetic chart of the basic divisions, some of the more dominant of the overlapping circles of a Venn diagram.

This chart contains a very limited view for the option of access to the future of organizing knowledge. The heart of the work is that mass of information, which includes those intermediaries in libraries and information centers, that recognize the dynamic processes of science and technology, the shifting values of interdisciplinary relations, and the complexity of the multidisciplinary problems of society such as energy and ecology. To put it another way, one cannot judge the contents of a book by the title on the cover.

We must all try to understand better the philosophy and principles of organizing information so that we have the basis for making intelligent decisions effectively. For example, it may well that DDC is a chemical and chemical engineering information service is especially rich at the broadest level of generality. For the information scientist that description is important. What are chemistry and chemical engineering?

Figure 3 lists the eighty services into which DDC subdivides chemistry and chemical engineering. Some of the eighty services are quite clearly chemical--such as Section 67, "Nonmetallic Inorganic Compounds (See Natural Gas)." Others, such as Section 66, "Glass Industries," exemplify the lack of definitive areas necessary to interdisciplinarity. Figure 4 lists the aspects of organic chemistry that we consider chemical. Figure 5 lists facets of organic chemistry that we classify into a different subfield, and it also lists figures that we feel fall outside of the scope of chemistry and, therefore, into another discipline. These decisions are arguable, of course, and we particularly agree that many overlap. The essential point is that we feel it necessary, in order to produce an information service that can be held to useful size and that can be used with assurance, to publish an operational definition of scope--one clear as possible a statement of our boundaries. The definition is almost ten hundred pages long--which indeed, of course. We are not telling the world that it should suffer to the same extent from any of the other specialized services; we are defining the parts that might otherwise and that we need to use effort to be comprehensive for the discipline. It is an information world that has information users, many information intermediaries, such as librarians, and other information specialists, and many policy makers such as in federal government agencies have not understood this situation or view enough ways to deal with it realistically.

The major Classification schedules have been working on the problem of definition of concepts for a long time. One of the ultimate objectives of their work is, of course, to stimulate automatic subject processing. "Subfreezing Agriculture" and "Herding" are virtual strikes to my business. Intensive Agriculture is automatic research is often necessary, even desirable. Subfreezing Agriculture is

Biochemistry Sections

1.	Pharmacodynamics	61653
2.	Hormone Pharmacology	61105
3.	Biochemical Interactions	61424
4.	Toxicology	61516
5.	Agrochemicals	61961
6.	General Biochemistry	62291
7.	Enzymes	62822
8.	Oxidation Biochemistry	62993
9.	Biochemical Methods	63091
10.	Microbial Biochemistry	63640
11.	Plant Biochemistry	63656
12.	Nonmammalian Biochemistry	63843
13.	Mammalian Biochemistry	63980
14.	Mammalian Pathological Biochemistry	64454
15.	Immunochemistry	64700
16.	Fermentations	64820
17.	Foods	64947
18.	Animal Nutrition	65213
19.	Fertilizers, Soils, and Plant Nutrition	65413
20.	History, Education, and Documentation	65661

Figure 3(1)

Organic Chemistry Sections

21.	General Organic Chemistry	65822
22.	Physical Organic Chemistry	65843
23.	Aliphatic Compounds	65875
24.	Alicyclic Compounds	66019
25.	Noncondensed Aromatic Compounds	66102
26.	Condensed Aromatic Compounds	66234
27.	Heterocyclic Compounds (One Hetero Atom)	66270
28.	Heterocyclic Compounds (More Than One Hetero Atom)	66466
29.	Organometallic and Organometalloidal Compounds	66786
30.	Terpenoids	66961
31.	Alkaloids	66989
32.	Steroids	66912
33.	Carbohydrates	66960
34.	Synthesis of Amino Acids, Peptides, and Proteins	67016

Figure 3(2)

Macromolecular Chemistry Sections

35.	Synthetic High Polymers	67083
36.	Plastics Manufacture and Processing	67334
37.	Plastics Fabrication and Uses	67709
38.	Elastomers, Including Natural Rubber	67870
39.	Textiles	67960
40.	Dyes, Fluorescent Whitening Agents, and Photographics	68237
41.	Leather and Related Materials	68290
42.	Coatings, Inks, and Related Products	68316
43.	Cellulose, Lignin, Paper, and Other Wood Products	68488
44.	Industrial Carbohydrates	68650
45.	Fats and Waxes	68677
46.	Surface Active Agents and Detergents	68830

Figure 3(3)

Applied Chemistry and Chemical Engineering Sections

47.	Apparatus and Plant Equipment	49675
48.	Unit Operations and Processes	49696
49.	Industrial Inorganic Chemicals	48807
50.	Propellants and Explosives	48910
51.	Fossil Fuels, Derivatives, and Related Products	48937
52.	Electrochemical, Radiational, and Thermal Energy Technology	49189
53.	Mineralogical and Geological Chemistry	49221
54.	Extractive Metallurgy	49560
55.	Ferrous Metals and Alloys	49765
56.	Nonferrous Metals and Alloys	50251
57.	Ceramics	50757
58.	Cement and Concrete Products	51008
59.	Air Pollution and Industrial Hygiene	51089
60.	Sewage and Wastes	51230
61.	Water	51408
62.	Essential Oils and Cosmetics	51604
63.	Pharmaceuticals	51632
64.	Pharmaceutical Analysis	51786

Figure 3(4)

Physical and Analytical Chemistry Sections

65.	General Physical Chemistry	51819
66.	Surface Chemistry and Colloids	52053
67.	Catalysis and Reaction Kinetics	52233
68.	Phase Equilibriums, Chemical Equilibriums, and Solutions	52333
69.	Thermodynamics, Thermochemistry, and Thermal Properties	52535
70.	Nuclear Phenomena	52905
71.	Nuclear Technology	53187
72.	Electrochemistry	53547
73.	Spectra by Absorption, Emission, Reflection, or Magnetic Resonance, and Other Optical Properties	53701
74.	Radiation Chemistry, Photochemistry, and Photographic Processes	54349
75.	Crystallization and Crystal Structure	54651
76.	Electric Phenomena	55092
77.	Magnetic Phenomena	55552
78.	Inorganic Chemicals and Reactions	55653
79.	Inorganic Analytical Chemistry	56043
80.	Organic Analytical Chemistry	56249

Figure 3(5)

wasteful, frustrating, and undesirable. It is probably completely preventable only in that best of all possible worlds, the future, when the ideal information system will be used in an ideal manner by ideal users.

As the accessing services strive to eliminate redundant processing, which they wish to do for the sake of their own efficiency, they must do so without making life more difficult and expensive for their users. Interdisciplinary fields like biochemistry and metallurgy grew, as the term indicates, between disciplines. So did--and do--the audiences for interdisciplinary information service. There are users between the discipline services as well as subject matter. Some biochemists have a biological orientation, some a chemical orientation. Both must be served, often with the same or similar information, just as they often generate the same or similar kinds of information. And there are many other examples. In the real world of producing useful information service there is no readily apparent easy solution to the problem of discipline overlap. Solutions there will be, but they are not readily apparent nor will they be easy.

Figure 6 gives a statistically based view of the distribution of subject matter through the scientific journal literature which provides about 70 percent of the documents we cite in CAS services. We select individual papers for analysis, based on their subject content as described in the 200-page subject coverage manual I mentioned. Between three hundred and four hundred journals currently provide 50 percent of the papers for our abstracts of selected journal papers. Less than three hundred journals are exclusively chemical. The last 15 percent of abstracts are of papers from journals predominately in other disciplines than chemistry, but they contain chemistry and that is what we are after.

The future of organizing knowledge depends upon a solution to the matter of access, and that solution must include new attacks on resource sharing among the major accessing services. In my view, it must also include recognition by government that the information services that it needs or that it feels the nation needs in energy, pollution, crime detection and other areas of societal concern must be built to the optimum extent upon the capabilities that existing services can bring to bear by resource sharing. Today, some are, many are not. Other approaches lead to waste, confusion, and much duplication of time and effort.

So far I have identified the major information transfer functions, and I have tried to illustrate the problem of subject matter dispersion. Now I should like to touch on some other aspects of the size and structure of the information store that we must make more accessible.

Figure 7 was borrowed from a recently published in-depth study by Dale Baker, Director of CAS. That figure describes the growth of the volume of chemical and chemical engineering literature. CA cited about 3.5 million documents between its founding in 1907 and 1965. Almost

SECTION 18
ANIMAL NUTRITION

A. Coverage in this Section includes:

1. General nutritional studies in all animal species, except protozoa (See Section 10), relating to vitamins, minerals, carbohydrates, lipids, proteins, and amino acids.
2. Evaluation of the nutritive value of foods and of nutritional balance and interrelations.
3. Animal requirements and utilization of vitamins, minerals, carbohydrates, lipids, proteins, and amino acids.
4. Energy values of food constituents (food calories), feed efficiency, and bioenergetics.
5. Diseases distinctly recognized as resulting from nutritional

Figure 4

SECTION 18
ANIMAL NUTRITION

B. Specific Exclusions and Alternative Placing

1. Intermediary metabolism of vitamins, minerals, and the other food elements: Section 12 (Nonmammalian Biochemistry); Section 13 (Mammalian Biochemistry).

⋮

4. Food composition unrelated to nutrition: Section 17 (Foods).

*5. Effects and utilization of chemically undefined foods in relation to growth, weight gain, or animal fertility: Excluded from cov- coverage in CA.

*6. Single chemical compounds fed with no chemical analytical effects: Excluded from coverage in CA.

*7. Clinical nutrition studies: Excluded from coverage in CA.

Figure 5

DISTRIBUTION OF CHEMISTRY IN THE SCIENTIFIC LITERATURE

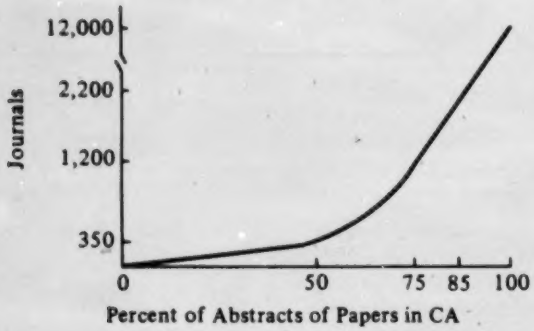
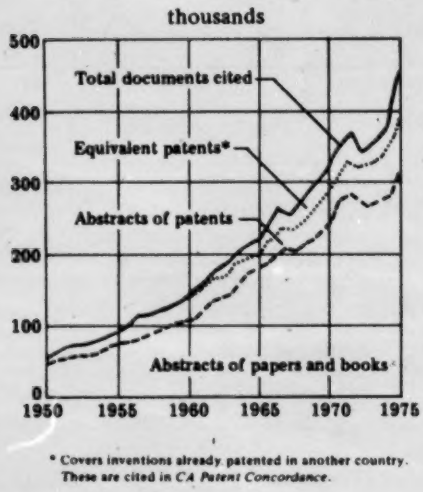


Figure 6

DOCUMENTS CITED IN CA,



* Covers inventions already patented in another country. These are cited in *CA Patent Concordance*.

Figure 7

WORLDWIDE NON - PATENT LITERATURE (1975)

National Origin		Pub. Language	
U.S.	25.8%	English	59.7%
U.S.S.R.	24.6%	Russian	23.3%
U.K.	6.3%	German	4.8%
Japan	7.3%	French	3.0%
Germany	6.8%	Japanese	3.0%
France	4.1%	All other	8.2%
All other	25.1%		

Figure 8

as many, 3.3 million, were cited between 1965 and 1975.

Figure 8 describes the internationality of the journal articles we abstract, and it summarizes the languages in which those papers are published. In a typical year, chemical science and technology will be conducted in 150 nations, and the original journal publications will appear in fifty languages.

In the ideal tomorrow, everything known will be available to everyone through a concatenation of systems and networks so powerful and so flexible that the user can find what he wants when he wants it even though he is unsure of what he is looking for when his search begins. New technology, effectively applied, is an essential component of that brave new world.

I am confident that progress in communications technology will be pursued and achieved. The community that is seeking new and better communications is large, diverse, and powerful. It encompasses much more than science information. But we need also to be sure that there is something in the network worth communicating, and so for the next few minutes, I will talk about an aspect of information technology that rarely captures public attention. That is the technology of creating the science information record so that it can be communicated.

In order to clarify how this type of information processing technology will influence the future of organizing knowledge, especially with respect to the matter of access to information, I will summarize present CAS production technology and contrast it with the traditional way of doing things. The National Science Foundation made a major contribution both in funds and moral support to the conversion of CAS from the old way to the new, and we express appreciation for that.

Until ten or twelve years ago, the technology used to create the record of man's scientific and technical achievements was only marginally different from that used when today's oldest primary journal began publishing in 1665. The focal point of that technology--its sole objective--was the creation of a printed document, and that objective remained fixed for many years, whether the document were a primary journal, an abstracting journal, a book, or some other form of publication.

Three hundred years after the first issue of the Philosophical Transactions of the Royal Society of London, Chemical Abstracts Service began to be one of the early users of newer technology. In the 1950s CAS staff began to dictate their abstracts and index entries into magnetic wire recorders. In 1961, Chemical Titles, the first periodical utilizing the Luhn-IBM Keyword-in-Context index programs, was produced on a borrowed computer. The first issue of an early if not the first computer-produced specialized abstracting service, Chemical-Biological Activities (CBAC), appeared in 1965, produced on the first CAS dedicated computer, with the first print chain to print upper and lower case characters, Greek letters, super and subscripts, and other concessions to human readers.

Within the CBAC development were the seeds of some concepts that were later to bloom impressively as the information processing revolution gathered momentum and broadened its objectives. One concept applied in CBAC was that indexes should be implicit in the original analysis of the primary document, in contrast to the traditional CAS approach which was to provide the indexer with an edited abstract as one of the raw materials of indexing. That CBAC concept has gone through radical evolution as I will point out.

Another quite major concept was the CAS Chemical Registry which was undergoing a feasibility test in 1965 and was first applied to CBAC. The Registry has now evolved into an automated vocabulary control and indexer support system, and it is the source, as well, of several huge and highly structured chemical information files that make possible new unique and powerful search capabilities.

Figure 9 has been used many times to clarify the conceptual difference between the traditional, pre-1966, CAS production system, and the system towards which we started moving ten years ago. At the time we began our systems development we felt that we could complete the conversion in eight to twelve years. The change would be revolutionary; the concepts and techniques had not been applied to publication; and we intended to maintain our quality and our production schedules without stopping or even delaying the publication of a single issue while the internal revolution went on. Planning and scheduling such a major development effort was a huge task. The conversion had to be stepwise, in hundreds of carefully coordinated steps, and critical path scheduling became and remains a way of life.

The meandering line at the left on Figure 9 characterizes the old manual system. The diamonds indicate essential human processing steps, and there were some thirty-five of them between receipt of a source document at the upper left square and completion of final indexes at the lower right. The stars were transcription, proofing, and correction steps. The process was hierarchical and highly redundant: it was based on progressive improvement of the recorded analysis of a document as it moved through the thirty-five steps.

The target system, to the right, is an integrated, on-line production system. The intellectual bucket brigade has been abolished.

We will not meet our original target of 1978 for completion of that system because in the absence of continued National Science Foundation (NSF) support we can not generate enough income as a not-for-profit organization to sustain the development program at its former rate. However, before NSF support ended, we were well along with many of the essential components of the ultimate system, and had started on-line staff support in editing functions. That particular application of on-line processing should be complete early in 1977.

THE CAS WORKFLOW

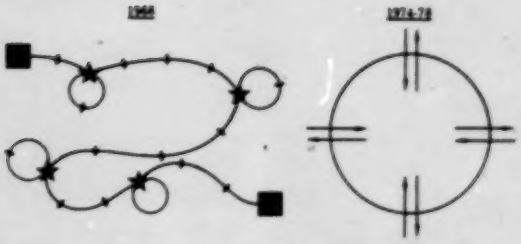


Figure 9

TARGET SYSTEM

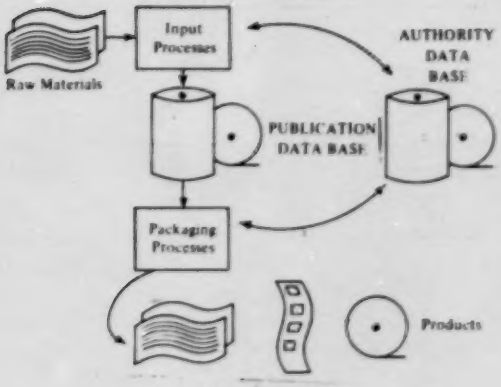


Figure 10

Obviously, a circle with pairs of arrows indicating interaction is not a completely descriptive diagram of our target system. Figure 10, still idealized to some extent, comes closer to characterizing the flow of our target production system. We cannot, after all, expect to build a one-step system. At the minimum there must be perhaps six steps to accomplish the same purpose as the original thirty-five.

In Figure 10, the raw materials are the "original literature" of Figure 1. At present, our input consists of primary journals, patents, conference proceedings, technical reports, edited monographs, and dissertations. In the future, some of these forms of original literature may evolve into non-existence, but new forms will probably come into being to serve the same basic functions which are unlikely to change very much. "Input processes" include selecting the specific documents to be analyzed, recording the identifying bibliographic information, and preparing and recording the abstract and two sets of subject index entries. One set of index entries is for the weekly printed issues of Chemical Abstracts and the computer-readable service, CA Condensates; the other set was originally conceived for the printed semiannual and collective indexes but is now also receiving broader use in biweekly computer-readable form. The input processes concentrate the intellectual work that was formerly strung throughout the system. All of the processes following input are automated including a variety of error detection routines and final output of services.

Two concepts--"data base" and "data element"--are probably most important to understanding how the CAS production system works and how it relates to improved user access.

The term "data base" is commonly used among information people with more than one meaning, both misleading. Information centers and on-line service organizations speak of the various "data bases" they have "up" for searching on their system. In Figure 10 "data base" is used in a different sense, to refer to a set of dynamic internal files which are the next-to-last step in the manufacture of CAS services. Some of us want to use the term "information base" for "data base" unless the file being referred to actually contains numerical or other measured information. I like the idea but came across it too late for the illustrations to this paper. Figure 10, therefore, shows the components of the CAS data base--the publications component and the "authority" component. The publications component contains the bibliographic information, the abstracts, and the subject index information that we derive from source documents. The authority component--with "authority" used purely in the documentation sense and usually misunderstood--provides vocabulary control and other types of processing support. The authority data base is also a source of search aids for users since it codifies editorial policy and practice, and extracts of these files are, by definition, guides to better understanding of our services.

"Data element" is by now a fairly common term, meaning a precisely and consistently defined unit of information with a specific content and format. All of the information processed by CAS is in the form of data elements.

In principle, a data element may be whatever a system designer and editor choose; they may wish to have a data element consist only of the last name of the senior author of a journal paper, or they may choose to clump the complete citation for a paper into one data element. But once defined and identified, a given data element must maintain that defined identity. The CAS data element dictionary lists hundreds of different data elements that may be used: the information being processed is finely divided and so permits great flexibility both in the design and assembly of services and in the use of those services when they are in computer-readable form.

It is useful, then, to consider the core of CAS to be the internal data base, or information base, which consists of a computer-readable array of discrete, separately addressable data elements. Outputs in the form of user services may consist of any desired subset of these data elements. All that is needed is a computer program to extract the desired elements and reassemble them into the desired output form. All of our printed publications, microforms, and computer-readable files are defined in terms of their data element content. Packaging programs are written in terms of those data elements. For publications that are to be printed, the content, the format, and the typography are defined in terms of data element identifiers. These identifiers trigger computer programs which select the content, format the page, including justification and hyphenation, and finally guide the computer-driven photocomposer in generating a graphic arts quality page image for the printer. That is today's system.

To illustrate, CA subject index entries may consist of up to fourteen data elements. Figure 11 is a list of those data elements and their codes. Figure 12 is an entry from the printed CA Formula Index, showing the data elements that are retrieved from the date base and placed in the desired order, all by computer program. A program also drives the photocomposer to set each element of the entry in the desired type face. Note that the name of the compound appears in light face type, indented and below the main entry point, which is the molecular formula.

Some of the same and some different data elements are needed for the printed Chemical Substance Index (Figure 13). The needed data elements are again retrieved from the data base, but their arrangement and typography are quite different. For example, this time the name is placed first in the entry, and it is set in a mixture of bold and light type to highlight the parts of the name. Since the Chemical Substance Index entry required a drawing of the ring system for that substance, the computer-driven photocomposer drew it also, and placed it in the proper position.

When that index information is needed for the computer-readable CA SUBJECT INDEX ALERT, which combines the contents of three CA subject matter indexes, typography is, of course, ignored, and the data elements are simply retrieved and placed where our "Standard Distribution Format" calls for them to be, each still separately addressable.

CAS SUBJECT MATTER INDEX DATA ELEMENTS

Concept Heading (CTH)	Name Modification (MOD)
Functional Category (CAT)	Stereochemistry (STE)
Qualifier (QUA)	Molecular Formula (FOR)
Heading Power (PHD)	Registry Number (RIN)
Monograph Definition (MON)	Text Modification (TMD)
Low Formula (LFM)	Primary Publication Type (PUB)
Substituent (SUB)	Abstract Number (CAN)

Figure 11

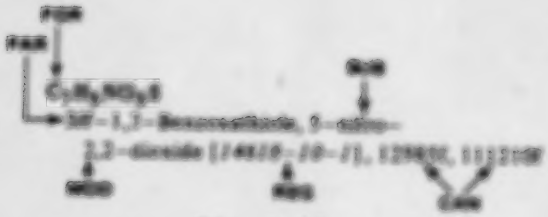


Figure 12

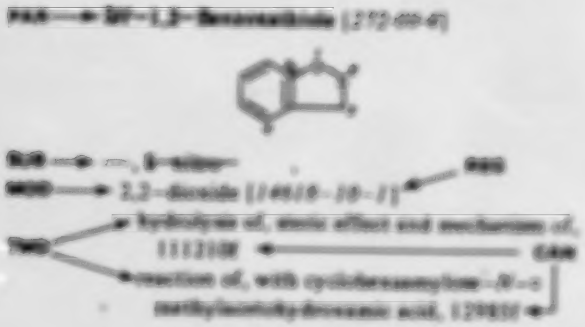


Figure 13

This has been a microcosmic view of the concept of single input, multiple outputs in action. One would hope that this capability, which extends to all the information we process, will entice the information-using public to break away from the traditional image of abstracting journal and index volume and to think more precisely in terms of the information mix they really need.

What are the implications of this technology for the future of organizing knowledge? Reduced to the simplest possible terms, one of the major mechanisms in the progress of mankind and society is the cycle of "information - knowledge - wisdom." The human race needs to take advantage of what it has learned. That part of the cycle that converts information to knowledge and knowledge to wisdom is outside the scope of my comments today, but I am deeply concerned with our weakness in the matter of access, whether it be to information, to knowledge, or to wisdom.

It is essential for the future of organizing knowledge that what mankind knows, as well as what we learn, be recorded in some accessible medium, form, and place and that effective means exist to facilitate access to the intellectual stockpile.

Part of the means will be new communications systems and hardware such as the personal computer terminal in every home that wants one. I am confident that the day is not far off when my son will buy for himself the terminal that he cannot now afford at its $1,900 price. The price may not fall as rapidly or as far as it has for pocket calculators and digital watches but fall it will.

But that cannot be the whole story. Other factors must also be right. David wants the terminal both for professional and for avocational use. For both purposes, he is fortunate to have access to ample computer power. For his job, there is a data base to work with. For his avocational interests, he is creating his own data base. The quality of both data bases is adequate to the intended use.

But what of the rest of mankind? If we assume, as I do, that an adequate personal terminal will in time be competitive in price with a good personal TV set or programmable pocket computer, what will the terminal connect with, and how? Will computer power and communications channels be as affordable as the terminal? And the data bases? What information will mankind want or need to access? Will it be there, and accessible?

In this age when cost-benefit analysis must be applied to virtually every move we make, and often at an elementary level, I despair of seeing the marvelous technology and the marvelously crafted information and data bases that we will be able to build at the service of the general public. Part of the problem, of course, is the difficulty of evaluating the benefits. The costs of information are easily calculated, but the benefits are not. It is easy to cost-justify million dollar payments to sports stars, because the financial data are easy to get. They

will probably create a greater return in efficiency with the amount invested in them. But what is the dollar benefit of an insight to a scientist, an engineer, or an ordinary citizen?

That is yet another philosophical challenge that I will walk away from. Instead, I will devote my concluding remarks to the simpler task of suggesting how attention to the nature of access will improve the future of organizing knowledge. Speaking only for CAS as one accessing service, there are several fundamental improvements that the user community is only beginning to exploit:

1. The data base concept in the production of information accessing service clearly makes for more efficient information processing. More efficiency means greater control over the production cost per unit of information.

2. The data element approach permits the greatest flexibility yet offered to the designer of information services. All of the content of the CAS data base is computer-addressable during processing. That means that literally any combination of information from the bibliographic, abstract text, and index areas can be exploited for service purposes.

3. Choice of output media to fit the type of service and the users' needs also greatly increases the potential search power of the user. In the CAS system, the medium, format, and other details of output do not reside in the data base but are determined only at output. Thus, the same service could be readily produced in printed, computer-readable, or microform, or any combination of these media, without reprocessing. Or, the particular characteristics of a medium can be applied to the special needs of a new service. Very large reference files that need frequent updating are usually economically unattractive printing projects. Graphic arts quality microfilm or fiche seem more attractive. We are planning to break out of the mold of the printed page and to try to create computer-readable files that are optimized for computer processing.

4. The authority file concept as practiced in the CAS system makes it possible to produce for the user community search support tools that make explicit the principles and techniques that are used in the production of the services to be searched.

5. The CAS system is not subject dependent. As we operate it, there are specifically chemical aspects, but for other applications those aspects need not be there. We are using part of our technology in cooperation both with

Bristol and also Information Science Services to provide the range of their services for small patrons. We operate a specialised information system for the National Cancer Institute and provide a significant part of the funding for some of the National Library of Medicine. We support of our publishing file formats [illegible] [illegible] files [illegible] to be efficient of files.

In summary, the ultimate objective of an information service is to provide access to information. The fundamental requirement is recognising the importance of this method of access and to the evolving capacities of all concerned members of society to apply our activities are and viable to support such cooperation. We may well be the best informed society in history, but we have much room for improvement. The past three years have seen much improvement and have identified paths for us to take. I am encouraged to believe we will make them.

THE MANY FACETS OF MARC

By

Lenore S. Maruyama

Unlike Christopher Robin who said "But now I am six, I'm as clever as clever. So I think I'll be six now for ever and ever," MARC has reached the grand old age of ten. An acronym for Machine-Readable Cataloging, MARC was first used as the name of a pilot project to create and distribute library cataloging records in machine-readable form. Since then, MARC has also been used in conjunction with several activities at the Library of Congress (LC), including a distribution service for machine-readable records, the formats for machine-readable records, and the units responsible for the design and implementation of projects to automate technical processing at LC and for the conversion of cataloging records to machine-readable form.

Each of these activities has had a significant impact on library automation and on outside libraries. Planning for the MARC pilot project, which was funded by the Council on Library Resources, was begun early in 1966. This effort was the direct result of three major studies and a series of conferences held to discuss the feasibility of converting cataloging records to machine-readable form (Refs. 1, 2, and 3). The Library of Congress selected a total of twenty libraries as participants in this pilot and from October, 1966 to June, 1968, distributed approximately 50,000 machine-readable records for English-language monographs.

The final report of this pilot stressed several points, many of which are still valid today (Ref. 4). Availability of high quality cataloging records from a central source in a timely fashion was deemed essential. The complexities of applying computer techniques to solve library problems were recognized by both the Library of Congress and the participating libraries, in addition to requirements for management support and for efficient management procedures to implement such projects.

Despite the difficulties, the results of the pilot project were sufficiently encouraging for the Library of Congress to continue the dissemination of machine-readable cataloging records on a regular basis. Beginning in March, 1969 with English-language monographs, the MARC distribution service operated by the Library's Cataloging Distribution Service has distributed approximately 656,000 records for monographic titles, 36,000 records for films, 30,000 records for serials, and 42,000 map records. Language coverage for the monographic titles is as follows: all current cataloging for English, French, German, Spanish, and Portuguese regardless of imprint date; and Dutch, Italian, Romanian, Norwegian, Swedish, Danish, and Finnish with imprint dates of 1976. Figure 1 depicts the present language coverage of the distribution services.

If funded, the Library expects to provide MARC records for all other titles in roman-alphabet languages and for music and sound recordings during fiscal year 1977. Figure 2 depicts the projected

MARC DISTRIBUTION SERVICE
PRESENT COVERAGE

FORM OF MATERIAL	STARTING DATE	DATES COVERED —	APPROXIMATE VOLUME PER YEAR
Books, English	Spring, 1969	1968-	80,000
Books, French	Summer, 1973	1973-	13,000
Books, German	January, 1975	1975-	21,000
Books, Spanish	January, 1975	1975-	8,500
Books, Portuguese	January, 1975	1975-	3,000
Books, Dutch/Scan	January, 1976	1976-	9,000
Books, Italian	January, 1976	1976-	7,500
Books, Romanian	January, 1976	1976-	2,500
Films	Fall, 1972	1972-	7,500
Serials	Spring, 1973	1973-	10,000
Maps	Spring, 1973	1973-	4,000

Figure 1

MARC DISTRIBUTION SERVICE
PROJECTED EXPANSION

FORM OF MATERIAL	STARTING DATE	DATES COVERED	APPROXIMATE VOLUME PER YEAR
Books, Other Roman Alphabet	January, 1977	1977-	10,000
Sound Recordings	January, 1977	1977-	6,000
Music	January, 1977	1977-	3,000
Cyrillic Alphabet	January, 1978	1978-	15,000
Other Non-Roman Alphabet	January, 1979	1979-	34,000
Name Authorities	January, 1977	1977-	120,000

Figure 2

MARC Users
(1975)

Primary Users		Secondary Users
Books	65	Estimated 1500
Films	15	
Serials	24	
Maps	14	

Notes: 15 commercial subscribers

22 outside U.S.

Figure 3

expansion. Distribution of authority records in machine-readable form will also begin with subject headings by November 1976. Figure 3 shows the number of MARC users.

An integral part of these distribution services has been the development of the machine record formats. A format had been developed for testing and evaluation during the MARC pilot project. This format, which became known as the MARC I format, had certain limitations, but it enabled the Library of Congress and the participating libraries to manipulate the records and the data for experimental purposes. Before the pilot project had been terminated, work had begun on the design of a new format.

Several factors had to be considered in the development of this new format, which became known as the MARC II format:

1. The format had to be hospitable for the transcription of bibliographic information for all kinds of materials.

2. The format had to be flexible enough to be used as the basis for automating all kinds of library activities in a wide range of libraries and institutions.

3. The design had to take into consideration a wide range of computer configurations and different kinds of programming languages which might be used in processing these machine records.

With such requirements, it was not possible to design a format that would be equally efficient for all users. It was felt that most institutions would have their own internal processing format to meet their particular needs, but to facilitate exchange and transmission of bibliographic data, it was necessary to have a standardized communications format. An institution would need only a single translation program to convert a local format to or from a communications format. Without a communications format, an institution receiving data from more than one source would need translation programs to convert the formats from each source to its own processing format. The MARC II format was designed to meet this requirement for communications purposes.

To date, the Library of Congress has developed the following MARC formats: books, films, maps, serials, manuscripts, and in process materials (Refs. 5, 6, 7, 8, 9, and 10). Formats for music and authorities are in press.

These MARC-related activities are performed principally by three units in the LC Processing Department: the MARC Editorial Division, the Cataloging Distribution Service, and the MARC Development Office.

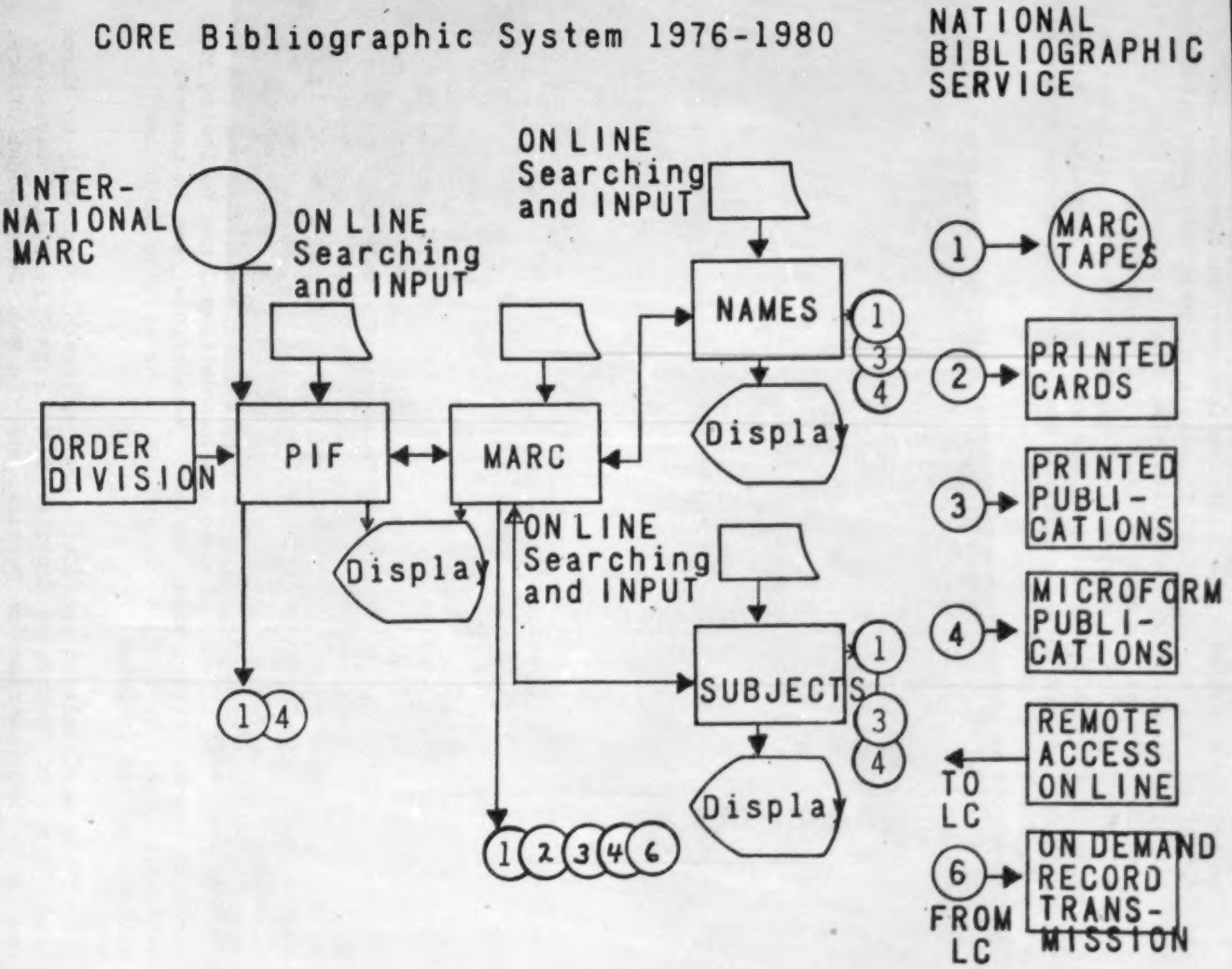

CORE Bibliographic System 1976-1980
NATIONAL BIBLIOGRAPHIC SERVICE
INTER-NATIONAL MARC
ON LINE Searching and INPUT
ON LINE Searching and INPUT
ON LINE Searching and INPUT
ON LINE Searching and INPUT
ORDER DIVISION
PIF
MARC
NAMES
SUBJECTS
Display
Display
Display
1 4
1 2 3 4 6
1
3
4
1
3
4
1 MARC TAPES
2 PRINTED CARDS
3 PRINTED PUBLI-CATIONS
4 MICROFORM PUBLI-CATIONS
REMOTE ACCESS ON LINE
TO LC
6 ON DEMAND RECORD TRANS-MISSION
FROM LC

Figure 4

The MARC Editorial Division has the primary responsibility for converting cataloging records for books and films and the name authority records to machine-readable form. (Conversion of serial, map, and subject authority records is handled by the Serial Record Division, the Geography and Map Division, and the Subject Cataloging Division, respectively.) The Cataloging Distribution Service is responsible for the production of a number of products, such as printed cards, book catalogs, and MARC tapes, from the machine-readable records. The MARC Development Office has been coordinating these activities as well as designing and implementing projects to automate technical processing in the Library of Congress.

The technical processing functions to be automated constitute the LC core bibliographic system, as depicted in Figure 4. A more complete description of these activities can be found in the publication The Library of Congress as the National Bibliographic Center so that only a brief discussion is given here (Ref. 11).

Acquisitions are handled by the Library Order Information System (LOIS) which presently operates as a separate system (see Figure 5). The acquisitions process at the Library of Congress is unique in that the bulk of materials is not obtained through purchase but through gift, exchange, copyright deposit, or transfer from other government agencies. Hence, only a relatively small number of records in the Order Division's machine-readable file represents new titles. The majority of items purchased have already been cataloged by the Library. LOIS provides the following capabilities: input and maintenance of the file; management reports and various printed products such as listings of the order file, purchase orders, etc.; status reporting as an item is processed within the Order Division; and fiscal control. It is expected that records for titles not already acquired and cataloged by the Library would be added to the online process file.

The process file is the first station where bibliographic control is applied to items received in the Library of Congress (see Figure 6). Searches are made in machine and manually-maintained files to determine if the item has not already been cataloged and if the heading to be used in the cataloging entry is established. Online input capabilities have been made available recently for a segment of the new items received, and these procedures will be expanded to other languages and types of materials as quickly as possible. Online status reporting procedures using bar code readers are also being implemented so that the location and status of an item and its corresponding cataloging copy can be monitored.

Work on authorities is in varying stages of development. Subject headings are in machine-readable form, and various products, such as printed lists, microforms, and distribution tapes, are being derived from this file. Input of name authorities is supposed to begin this

LOIS

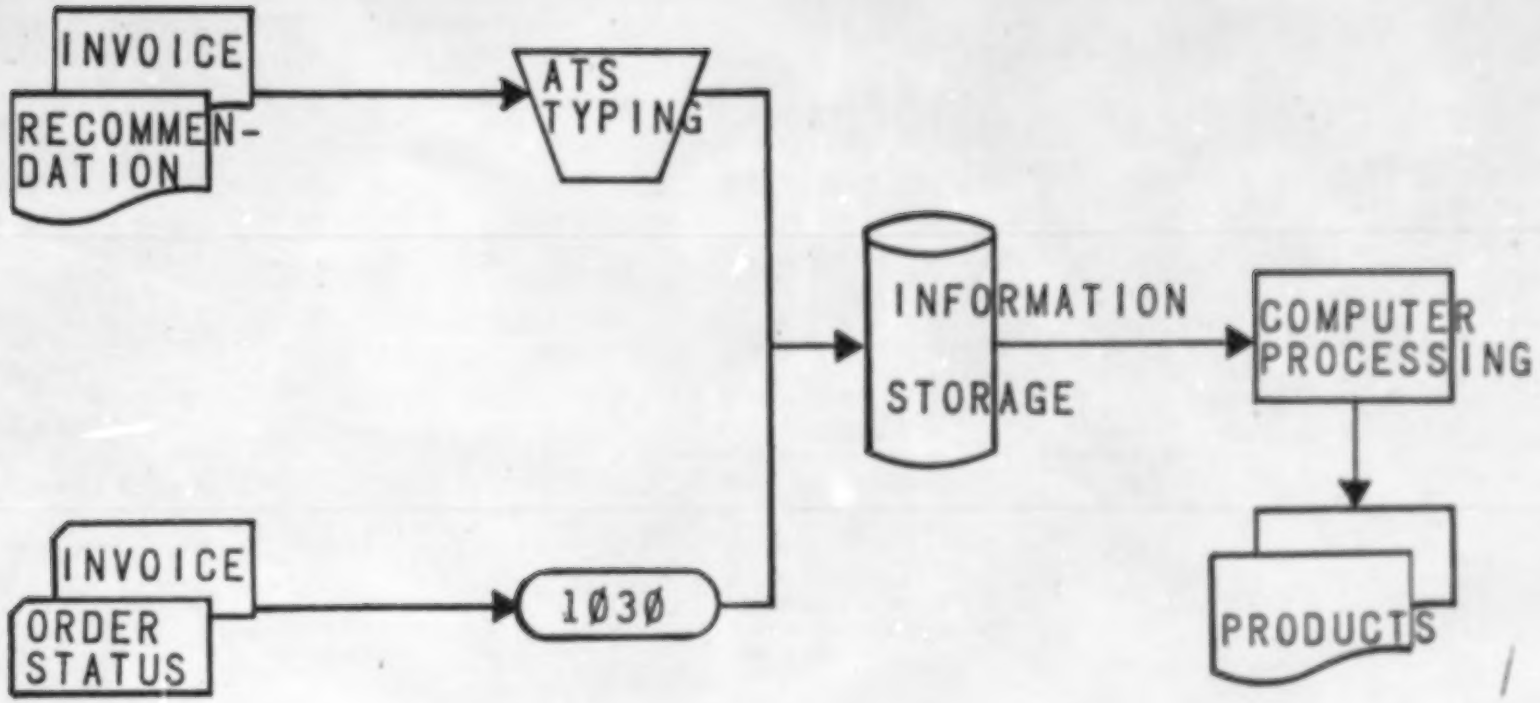


Figure 5

AUTOMATED PROCESS INFORMATION FILE LOCATION REPORTING SYSTEM

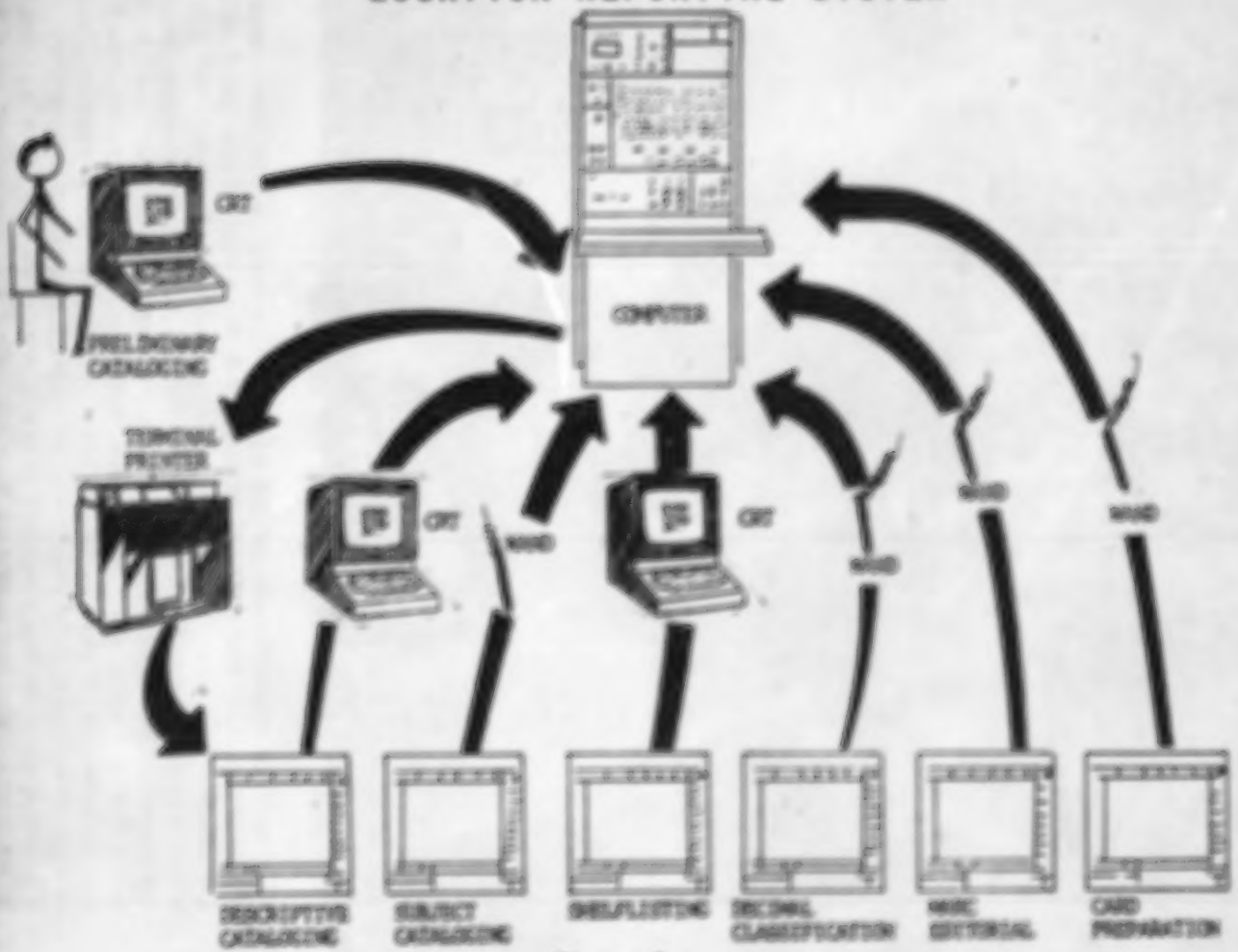


Figure 6

fall with the hiring of additional staff in the MARC Editorial Division. Further work on making these files available for online retrieval and integrating the authority and bibliographic files is in abeyance pending a reorganization within the MARC Development Office and a readjustment of priorities by the Processing Department. The MARC bibliographic record files, together with the authority files, constitute one of the most important modules of the core bibliographic system. True online cataloging cannot take place until these two components have been linked.

The products derived from the different components of the core bibliographic system then become part of the national bibliographic service provided by the Library of Congress to the nation's libraries. Present products derived from machine-readable files include printed cards, book catalogs, microforms, and magnetic tapes. Online access to LC files is planned as a new service, which will be described in more detail in a later section.

The impact of all of these components of MARC has been significant in the areas of standardization, including formats, character sets, hardware, and shared cataloging systems. The MARC format structure has been adopted by the American National Standards Institute and the International Standards Organization for interchange of bibliographic information on magnetic tape. The LC MARC formats have also been the basis for the development of MARC formats in Great Britain, Canada, Australia, Japan, etc., as shown in Figure 7. In order to facilitate exchange of records between these systems, work was begun and is nearing completion on an international MARC format or UNIMARC. Since each national system was developed to meet its unique requirements, the UNIMARC format would allow each national system to maintain its own format but would provide a standard format for international exchange purposes. An individual national bibliographic agency, e.g., the Library of Congress in the United States, would be responsible for converting data to the standard format of its country and translating its format to the UNIMARC format for exchange purposes.

Another significant impact of MARC has been on the development of character sets. The Library of Congress was responsible for compiling a character set that would allow the transcription of data in all roman alphabet languages and the romanized forms of non-roman alphabet languages. This standard library character set was then used as the basis for defining the appropriate codes in the American Standard Code for Information Interchange (ASCII). Since the ASCII standard consisted of a 7-bit code which was not large enough to accommodate the 171 characters in the standard library set, the ASCII set was expanded to an 8-bit code for users of 9-track tapes and contracted to a 6-bit code for users of 7-track tapes. The Library of Congress, through its MARC distribution, was the first agency to use the ASCII standard. At present, work is proceeding on defining standards for Latin, Greek,

INTERNATIONAL MARC

PLANNED OR OPERATIONAL	UNDERSTUDY
Australia	Poland
Belgium	Soviet Union
Canada	
Denmark	
France	
Italy	
Japan	
Latin America	
Norway	
South Africa	
Spain	
United Kingdom	

Figure 7

and Services Division are to collaborate with the International Programs Department.

The definition of a standard library character set also has the support of hardware. During the last seven years, the existence of ASCII has led to the printing capabilities of the Library Processing Service, photocomposition device, CRT device, and terminal printer, as well as input and display capabilities of at least three different CRT terminals.

MARC's impact on libraries and library automation has been significant. The availability of MARC tapes at such a reasonable cost has enabled a large number of commercial firms and service bureaus to market a wide range of profitable products. Because of high costs, relatively few libraries are processing their data on systems, but a large number of them are sending tapes to firms specializing in various types. The Ohio Bibliographic Network, of which Kent State University Libraries is a part, is an example. The newly-formed Network Development Office at the Library of Congress has been established to coordinate and monitor plans, programs, and budgets to support automated systems related to the full range of bibliographic and technical services that support the future national library network. An important aspect of this office's work is to work in liaison with the regional networks and service centers to ensure that the components of the national network will fit together. We are in the process of compiling a position paper, with the assistance of an advisory group, on the functions to be performed by the network, its possible configuration, and its membership. This paper will be given wide dissemination to solicit comments from libraries in the field.

In addition to the work on long-range plans, this office is involved in several on-going projects that constitute some of the present work for the national library network. A pilot project is in progress to provide a computer-to-computer connection between the Library of Congress and the Research Libraries Group's facility at the Red Mill Public Library. Several studies are in progress or will be started to investigate the role of an automated file in a national network, to determine the role of the Library of Congress in this network, to conduct a survey of machine-readable data bases from which location reports for a national union list can be obtained, and to provide several directions to the projects involving "shared" records, tapes and CRTs.

The key to the future is the library community, including the Library of Congress, in "sharing." Such a goal will ensure that limited resources will be expended on projects that will help many institutions instead of just one and that the ultimate users of the bibliographic resources in libraries will be served in a timely and efficient manner.

REFERENCES

1. King, Gilbert W. and others. Automation and the Library of Congress; A Survey Sponsored by the Council on Library Resources, Inc. (Washington, Library of Congress, 1963) 88 p.

2. Buckland, Lawrence F. The Recording of Library of Congress Bibliographical Data in Machine Form; A Report Prepared for the Council on Library Resources, Inc. Rev. (Washington, Council on Library Resources, 1965) 54 p. (Photocopy available from the LC Photoduplication Service, Washington, D.C. 20540. $8.00 postpaid)

3. Avram, Henriette D., Freitag, Ruth S., and Guiles, Kay D. A Proposed Format for a Standardized Machine-Readable Catalog Record; A Preliminary Draft. June 1965. Reprinted with index and appendix. (Washington, Library of Congress, 1971) 136 p. ISS Planning Memorandum no. 3.

4. Avram, Henriette D. The MARC Pilot Project; Final Report on a Project Sponsored by the Council on Library Resources, Inc. (Washington, Library of Congress, 1968) 183 p. (For sale by the Supt. of Docs., U.S. Govt. Print. Off., Washington, D.C. 20402. $5.15. LC 1.2:M18/2. S/N 3000-0009)

5. United States. Library of Congress. MARC Development Office. Books: A MARC Format; Specifications for Magnetic Tapes Containing Catalog Records for Books. 5th ed. (Washington, Library of Congress, 1972) 106 p. (For sale by the Supt. of Docs., U.S. Govt. Print. Off., Washington, D.C. 20402. $1.75. LC1.2:M18/11. S/N 003-000-00057-9)

6. United States. Library of Congress. MARC Development Office. Films: A MARC Format; Specifications for Magnetic Tapes Containing Catalog Records for Motion Pictures, Filmstrips, and Other Pictorial Media Intended for Projection. (Washington, Library of Congress, 1971) 65 p. (For sale by the Supt. of Docs., U.S. Govt. Print. Off., Washington, D.C. 20402. $1.05. LC 1.2:M18/8. S/N 3000-00048)

7. United States. Library of Congress. MARC Development Office. Maps: A MARC Format; Specifications for Magnetic Tapes Containing Catalog Records for Maps. 2d ed. (Washington, Library of Congress, 1976) 76 p. (For sale by the Supt. of Docs., U.S. Govt. Print. Off., Washington, D.C. 20402. $1.40. LC1.2:M18/6/976. S/N 030-000-00079-0)

8. United States. Library of Congress. MARC Development Office. Serials: A MARC Format; Specifications for Magnetic Tapes Containing Catalog Records for Serials. 2d ed. (Washington, Library of Congress, 1974) 104 p. (For sale by the Supt. of Docs., U.S. Govt. Print. Off., Washington, D.C. 20402 $1.50. LC 1.2:M18/14. S/N 3000-00071)

9. United States. Library of Congress. MARC Development Office. Manuscripts: A MARC Format; Specifications for Magnetic Tapes Containing Catalog Records for Single Manuscripts or Manuscript Collections. (Washington, Library of Congress, 1973) (For sale by the Supt. of Docs., U.S. Govt. Print. Off., Washington, D.C. 20402. $.90. LC1.2:M18/12. S/N 3000-00063)

10. United States. Library of Congress. MARC Development Office. In Process Materials: A MARC Format; Specifications for Magnetic Tapes Containing Records for Materials in the Process of Being Cataloged in the Library of Congress. (Washington, Library of Congress, 1975) 29 p. (For sale by the Supt. of Docs., U.S. Govt. Print. Off., Washington, D.C. 20402. $.85. LC 1.2:M/15. S/N 030-000-0080-3)

11. The Library of Congress as the National Bibliographic Center; Report of a Program Sponsored by the Association of Research Libraries, October 16, 1975. (Washington, Association of Research Libraries, 1976) 58 p.